DOCUMENT RESUME

ED 254 541 TM 850 137

AUTHOR	Cooper, Harris M.
TITLE	A Taxonomy of Literature Reviews.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	Mar 85
GRANT	NIE-G-82-0022
NOTE	49p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985) under the title "The Literature Review: Knowledge Synthesis Activities in Education and Psychology."
PUB TYPE	Speeches/Conference Papers (150) -- Information Analyses (070) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Classification; *Content Analysis; Databases; *Educational Research; *Evaluation Criteria; Information Utilization; Interrater Reliability; *Literature Reviews; *Psychological Studies; Research; Surveys; Synthesis
IDENTIFIERS	ERIC; Psychological Abstracts

ABSTRACT

A taxonomy for literature reviews in education and psychology is presented. The increased use of the descriptor "literature review" in ERIC and Psychological Abstracts documents between 1969 and 1983 is cited as creating the need for categorization. The taxonomy categorizes reviews according to focus, goal, perspective, coverage, organization, and audience. The seven winners of the American Educational Research Association's Research Review Award are used to illustrate these categories. Data on intercoder reliability of taxonomy codings when applied by readers is presented. The taxonomy is used to describe a representative sample of existent reviews. Suggestions are made concerning how the taxonomy might facilitate judgments concerning the quality of future knowledge synthesis activities. General standards for evaluating reviews are presented. (DWH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED254541

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

A Taxonomy of Literature Reviews

Harris M. Cooper
Center for Research in Social Behavior
111 E. Stewart Road
University of Missouri-Columbia
Columbia, MO 65211

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. Cooper
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This paper was presented as an Invited Address to the annual meeting of the American Educational Research Association, Chicago, 1985, under the title "The Literature Review: Knowledge Synthesis Activities in Education and Psychology." Preparation of the paper was supported by National Institute of Education grant #NIE-G-82-0022, though the opinions expressed do not necessarily reflect those of NIE. The author wishes to thank Lee Shulman, Ron Ribble, and David Tom for their help throughout the project.

TM 850 137

Running head: Literature Reviews

Abstract

A taxonomy of literature reviews in education and psychology is presented. The taxonomy categorizes reviews according to: (a) focus; (b) goal; (c) perspective; (d) coverage; (e) organization; and (f) audience. The seven winners of the American Educational Research Association's Research Review Award are used to illustrate the taxonomy's categories. Data on the reliability of taxonomy codings when applied by readers is presented. Results of a survey of review authors provides baseline data on how frequently different types of reviews appear in the education and psychology literature. How the taxonomy might help in judging the quality of literature reviews is discussed, along with more general standards for evaluating reviews.

The two major abstracting services in psychology and education, the Educational Resources Information Center (ERIC) and Psychological Abstracts, use the term "literature review", or a variation thereof, to describe the documents they contain. Figure 1 presents the percentage of all documents in each database that were assigned the term "literature review" between the years 1969 and 1983. The trend for both databases is for an increasing percentage of documents to be assigned this descriptor over the period covered, with an exceptional expansion in the use of the term by Psychological Abstracts between 1979 and 1983.

Place Figure 1 about here

The probable cause for the growing prominence of reviews in the education and psychology literature is the increased numbers of personnel and the accompanying information explosion that has occurred in these disciplines. According to Garvey and Griffith (1971): ". . . the individual scientist is . . . overloaded with scientific information and [can] no longer keep up with and assimilate all the information being produced that [is] related to his primary specialty" (p. 350). The response to this overload appears to be, first, a narrowing of specializations in which social scientists attempt to keep up with primary research and scholarship, and second, a greater reliance on literature reviews to remain abreast of developments in other fields of interest. Also, regardless of the assimilating capacities of social scientists, expanding literatures necessitate the collecting, evaluating, and synthesizing of scholarship in order to bring coherence and perspective to problem areas.

The enhanced role of the literature review requires that this expository form be given careful scrutiny. To date, such examinations have been scarce. The only aspect of literature reviewing that has received prolonged attention is the integration of empirical research. This concern is primarily an outgrowth of the introduction of meta-analysis procedures (Glass, McGaw & Smith, 1981; Rosenthal, 1984). However, a survey of recent authors of literature reviews, to be discussed momentarily, reveals that less than one-fifth of all reviews are undertaken for the purpose of exhaustively synthesizing a research literature. The majority of reviews are conducted for other purposes, and these synthesis activities have been almost completely neglected.

In this paper, I will attempt to correct this omission by offering a general definition of the term "literature review" and a taxonomy for classifying literature reviews according to their major characteristics. I will then illustrate the taxonomy by applying it to the past winners of AERA's Research Review Award. Some data on intercoder reliability will be presented and the taxonomy will be used to describe a representative sample of existent reviews. Finally, I will make some suggestions concerning how the taxonomy might facilitate judgments concerning the quality of future knowledge synthesis activities.

Because the existing literature on literature reviews hardly forms the basis for a review itself, I have supplemented the prior works on this topic in two ways. First, I conducted in-depth, unstructured interviews with fourteen scholars in diverse fields of education and psychology who were conducting literature reviews. The interviews occurred at several points during the reviewing process and touched on all aspects of the task, from problem formulation to editorial remarks. Second, based on the

interviews and on input from numerous people involved in the generation of knowledge syntheses, including the directors of ERIC clearinghouses and the National Institute of Educations' Dissemination in Practice program staff, a structured questionnaire was developed and completed by 68 scholars who had recently published reviews of research literatures (see Cooper, 1984a). Several of the results of this survey aided in the formulations I will present.

Of course, this paper is not intended to be a definitive statement on the nature of literature reviews. Instead, it is to be a working document meant to stimulate future discussion of the goals, processes, and problems associated with the literature review and, by implication, to help practicing literature reviewers produce documents of maximum utility for their audiences.

A Definition of the Literature Review

I decided to begin the search for a definition of the term literature review by examining the definitions used by ERIC and Psychological Abstracts. In the Thesaurus of ERIC Descriptors, the descriptor term "literature review" was accompanied by the scope note "surveys of the materials published on a topic" (ERIC, 1982, p. 143). The ERIC Processing Manual (Section 5: Cataloging: ERIC, 1982) contained the following definition for the literature review as a document type: "Information analysis and synthesis, focusing on findings and not simply bibliographic citations. Summarizing the substance of the literature and drawing conclusions from it" (p. 85).

The Thesaurus of Psychological Index Terms (American Psychological Association, 1982) provides no definitions for the document types it assigns in cataloging the literature. An inquiry to the offices of

Psychological Abstracts revealed that the document term "literature review" had no specific or formal definition. Instead, the definition of the term and its appropriateness for a document was left to the intuitive judgment of the indexer, with the proviso that the document had to be exclusively or primarily a literature review (that is, not also contain a report of primary data) for the term to be employed (D. Langenberg, personal communication, March 14, 1984). As one lexicographer at ERIC put it: "You knows one when you sees one." The scope note for the descriptor "literature review" in APA's Thesaurus defined these documents as "surveys of previously published material" (APA, 1982, p. 96) and also stipulated that the document should be entirely or primarily a literature review for the descriptor to be applied.

Another potential source of definitions was journals that specialize in publishing literature reviews. To this end, the policy statements of the Review of Educational Research and the Psychological Bulletin were examined. The Review of Educational Research policy statement says that the journal "contains integrative reviews and interpretations and educational research literatures on both substantive and methodological issues." Psychological Bulletin's policy states the journal publishes "evaluative and integrative reviews and interpretations of substantive and methodological issues in scientific psychology." Further, "Integrative reviews that summarize a literature may set forth major developments within a particular research area, or provide a bridge between related specialized fields . . .". Finally, original theoretical statements that contain literature reviews are not considered the province of Psychological Bulletin, but literature reviews that "develop an integrative theoretical statement" are acceptable.

It seems clear that a general definition of a literature review must contain at least two elements: First, a literature review uses as its database reports of primary or original scholarship, and does not report new primary scholarship itself. The primary reports used in a literature review may be verbal, but in the vast majority of cases reports are written documents. The types of scholarship may be empirical, theoretical, critical/analytic, or methodological in nature.

Second, a literature review seeks to describe, summarize, evaluate, clarify, and/or integrate the content of the primary reports. This second part of the definition implies that literature reviews are generally inductive in nature, a quality made explicit in the Psychological Bulletin's definition. However, the relation between existing theories and literature reviews is not that simple. For instance, sometimes the documents being evaluated and integrated in a literature review are themselves theoretical statements or other literature reviews. Other times, theoretical positions form the framework for evaluation and integration, thus rendering the review more hypothetico-deductive in character. This issue leads away from the problem of how generally to define the form to the problem of how to distinguish among different types of literature reviews.

Types of Literature Review

Previous attempts at defining types of literature review have primarily been concerned with the foci and goals of reviews, with particular attention paid to reviews that summarize empirical research. For instance, Jackson (1980) offered the following set of goals:

> Some (reviewers) are primarily interested in sizing up new substantive and/or methodological developments in a given field. Some are primarily interested in verifying existing theories or developing new ones. Some are interested in synthesizing knowledge from different lines of research, and still others are primarily interested in inferring generalizations about substantive issues from a set of studies directly bearing on those issues (p. 438).

To this list might be added Taveggia's (1974) notion that reviews are meant to highlight important issues that research has left unresolved, and Price's (1965) notion that reviews are meant to replace papers that have fallen behind the research front.

These definitions provide a starting point for a more exhaustive taxonomy. They highlight some of the central foci and goals of reviews. Because they deal primarily with integrative research reviews, however, they do not capture many varying aspects of the documents that fall within the definition of literature review provided earlier. Therefore, I would like to systematize and expand on these foci and goals and also suggest several other characteristics that usefully distinguish among literature reviews. These include: the perspective of the reviewer; the intended coverage of the review; the organization of the review; and the review's intended audience. Table 1 presents the six characteristics and their related categories. I will describe each briefly, then demonstrate how the taxonomy can be applied.

Place Table 1 about here

Focus. The focus of a review concerns the material that is of central interest to the reviewer. Most reviews in education and psychology center on one or more of four areas: research outcomes, research methods, theories, and practices or applications. The four foci are self explanatory and familiar to social scientists. Of course, they are not mutually exclusive areas of interest; in fact, it is rare for a review to have only a single focus. Instead, most reviews will have two or three foci that are given varying degrees of attention.

Goals. The second characteristic of a review is its goals. Goals concern what the author hopes the review will accomplish. The most obvious goal for a review is to integrate or synthesize past literature that is believed to relate to the same issue. In fact, this goal is so pervasive among reviews that it is difficult to find reviews that don't attempt to synthesize works at some level.

In their article on types of synthesis, Strike and Posner (1983) identified numerous activities that could be counted as integrative and that are often performed by literature reviewers. These include (a) formulating general statements from multiple specific instances, a type of synthesis common in research reviews, (b) resolving the conflict between contradictory ideas or statements of fact by proposing a new conception that accounts for the inconsistency, and (c) bridging the gap between theories or disciplines by creating a common linguistic framework.

While synthesis is pervasive among literature reviews, reviews can have other goals. For instance, reviewers may write for the purpose of critically analyzing the existing literature. Many reviews are judgmental about the work they focus on, be it research, theory, or practice. The intention of these reviews is usually to demonstrate that past conclusions

derived from the literature were unwarranted. The conclusion of unworthiness is typically based on the literature's incommensurability with the reviewers' theoretical stance and/or criteria for methodological validity. Unlike a synthesis, a review that concentrates on criticism less often compares the covered literature one to another but instead holds each instance up against a criterion and finds it either acceptable or not.

A third goal that is often at the heart of reviews is to identify issues central to a field. These issues may involve (a) questions that have dominated past endeavors, (b) questions that should dominate future endeavors, or (c) methodological problems that have prevented a topic area from progressing. While reviews emphasizing central issues usually provide suggestions about how problems and controversies in an area might be overcome, they are not necessarily syntheses because they don't always formulate generalities, attempt to resolve conflict, or suggest bridges between areas. However, as with foci, reviews more often than not have multiple goals. Frequently, integration and criticism or integration and identification of central issues go hand in hand.

Perspective. A third characteristic that distinguishes reviews concerns the point of view the reviewer employs in discussing the literature. Two perspectives can be identified: (a) neutral or dispassionate representation and (b) espousal or advocacy of a position or perspective. In the former, the reviewer tries to distill the works in a topic area with as little personal interpretation and evaluation as possible. The attention given to different theories, methods, issues, or outcomes is meant to reflect their relative prominence in the pertinent literature. In essence, the reviewer attempts to play the role of an "honest broker."

The second perspective is the opposite of the first. Here, the reviewer deliberately plays the role of an advocate and espouses the virtues of a particular paradigm, theory, methodology, or practice. The reviewer undertakes the task of accumulating and synthesizing the literature in the service of demonstrating the value of a particular point of view.

While Table 1 presents the two perspectives as separate approaches, it is probably best to think of them as opposite ends of a continuum. We might label the continuum "degree of active construction" by the reviewer. Whether reviewers can in fact be "honest brokers" is a debate receiving considerable attention among philosophers of science, and the arguments need not be reiterated here (see Philips, 1983; Eisner, 1983). However, it is important to note that while reviewers cannot maintain both the descriptive and interpretive perspectives simultaneously, they can switch modes within the same work, first describing dispassionately the contents of an area and then applying a particular perspective to it. We shall see, however, that such approaches are rare--most authors opt for one end of the continuum or the other and maintain that stance throughout the review.

Coverage. The next characteristic, coverage, is probably the most distinct aspect of literature reviewing. The extent to which reviewers find and include relevant works in their paper is the single activity that sets this expository form apart form all others. How reviewers search the literature and how they make decisions about the suitability and quality of material involve methods and analytic processes that are unique to this form of scholarship (see Cooper, 1984a).

The typology distinguishes between four types of coverage. The first level, exhaustive coverage, means the reviewer intends to be comprehensive

in the presentation of works relevant to the topic under consideration. An effort is made to include the entire literature or most of it, not just a sample, and to base conclusions and discussions on this all-inclusive information base. In this type of paper, the author describes all the works relevant to the conclusions that are drawn, but perhaps not in great detail.

The second type of coverage also bases conclusions on entire literatures but only a selected sample of works are actually described in the paper. The strategy for selecting works to cite might follow either of the patterns to be described momentarily. Especially in research integrations, authors often formulate conclusions in very general terms, using, for instance, phrases like "In summary, the research indicates . . ." or "The literature on this topic reveals . . .". Such statements imply a comprehensive coverage, but not necessarily that the work cited in the text exhausts the literature.

From the reader's perspective the distinction between exhaustive coverage and exhaustive coverage with selective citation is important. A reviewer who presents the entire information base allows the reader to evaluate (a) whether the coverage was, in fact, exhaustive and (b) whether the conclusions are warranted by the works included. The reviewer who has drawn general conclusions but only cites selected works (or makes no claim concerning how cited material was chosen) does not allow the reader to perform such an evaluation.

Some reviewers will opt for a third coverage strategy--presenting works that are representative of many other works in a field. A sample is presented that typifies larger groups of material. The author discusses the characteristics that make the sample illustrative of the larger group.

In this strategy, the author freely chooses the particular works that are deemed representative but the classes of material that need to be attended to are really not within the reviewer's discretion. Instead, this is a function of the frequency with which works that share particular characteristics appear in the literature.

In the final coverage strategy, the reviewer concentrates on works that have been central or pivotal to a topic area. This may include materials that initiated a line of investigation or thinking, that changed how questions were framed, that introduced new methods, that engendered important debate, or that performed a heuristic function for other scholars. Rather than being representative, a review that covers pivotal works describes important initial efforts that have provided direction for a field.

As with the previous characteristics, a particular review can employ more than one coverage strategy. Obviously, the exhaustive and exhaustive/selective strategies are mutually exclusive, at least within the same topic domain. However, it may not be uncommon for the representative and pivotal strategies to occur together.

Organization. How a paper is organized is a fifth characteristic that differentiates research reviews. Reviews can be arranged (a) historically, so that topics are introduced in the chronological order in which they appeared in the literature, (b) conceptually, so that works relating to the same abstract ideas appear together, or (c) methodologically, so that works employing similar methods are grouped as subtopics. Reviews can combine organizations by, for example, addressing works historically within a given conceptual or methodological framework.

Audience. Finally, the intended audiences of reviews can differ from one another. Reviews can be written for groups of specialized researchers, general researchers, practitioners, policy makers, or the general public. The audience distinction probably manifests itself most clearly through the writing style of the reviewer. As reviewers move from addressing specialized researchers to addressing the general public, they employ less jargon and detail while often paying greater attention to the implications of the work being covered. Of course, it is rare to find literature reviews that speak directly to the general public. Instead, reviews written for more specialized audiences are sometimes distilled and simplified by popular writers before appearing in periodicals intended for large general audiences.

Applying the Taxonomy to Award-Winning Reviews

In order to illustrate how the taxonomy can be applied, and to uncover problems in its application, my two research assistants, David Tom and Ron Ribble, and I undertook the pleasant task of reading the seven reviews that have won AERA's Research Review Award. We independently attempted to describe each of the reviews by using the characteristics and categories in Table 1. Table 2 presents the fruit of our labor. Contained in each cell are those categories that at least two of us agreed pertained to the review. Half of the listed categories received three votes and half received two votes. One of every eight category nominations received only one vote. These are not listed. Before examining the table, it will be instructive to detail how the taxonomy was applied to one of the reviews.

Place Table 2 about here

Noreen Webb (1982), the 1984 winner of the Research Review Award, performed a review concerning student interaction in small learning groups. Webb's focus was to "examine research bearing on the relationship between interaction and achievement and research exploring the predictors of interaction in small groups" (p. 422). At the end of the paper, some, but considerably less, attention was given to research methods through Webb's discussion of interpretive problems arising from "noncomparable designs, lack of detailed or appropriate observation procedures, inappropriate unit of observation and simplistic analytic strategy" (p. 439).

The integration goal of Webb's review is exemplified by her use of summary statements such as "the research relating interaction in groups and achievement generally shows that giving help and receiving help are positively related to achievement, and off-task and passive behavior are negatively related to achievement" (p. 427). At the end of her paper, she identifies central issues in the area, based primarily on her assessment of research design and observational problems with past research.

Though she clearly believes interaction variables can enhance our understanding of small group learning, Webb takes a generally neutral position toward the research meant to demonstrate this contention. In one instance she calls the research "not sufficiently consistent at this time to warrant an unqualified conclusion" (p. 441).

One problem that arises in applying the taxonomy is illustrated by our approach to the first three categories. Taxonomy users are faced with the decision of whether to apply the categories from the perspective of a reader or from the inferred perspective of the author. In some instances the category nominations might differ. Thus, a reviewer might claim neutrality toward an area but a reader might perceive the paper is

an example of advocacy. In our application, we attempted to infer the intent of the author when making our judgments. Either approach can be employed, however, and an interesting set of issues arise concerning disparities between an author's expressed intentions and what they accomplish. These issues relate to judgments of review quality, to be discussed later.

With regard to coverage, Webb attempted to be fairly exhaustive, within the limiting criteria of only including studies involving individual learning and systematically measured interaction (pp. 422-423). Applying the coverage categories led us to consider whether a literature review can ever be truly exhaustive. All authors of reviews must necessarily exclude a multitude of work that lies near the boundary of their problem domain, works that other reviewers might choose to include. To solve this problem in applying the label, we chose to operationally define "exhaustive" as meaning comprehensive coverage within the limitations of the author's definition of the area. We also chose to label as exhaustive reviews that confined themselves to particular time periods, for example all research conducted after 1975, if the author comprehensively examined the delineated period. Other users of the taxonomy might choose to operationally define exhaustiveness in a different manner.

Webb's review was organized by grouping studies that shared the same conceptual underpinning, though the concepts might be termed narrowly abstract. For example, her categorization of research under headings such as "helping behavior", "off-task and passive behavior", "ability groups composition and reward structure" are concepts closely tied to observable measurement procedures. In discerning an author's organization scheme, we

found a good indicator was the headings employed to distinguish subtopics within the paper.

Finally, the level of specificity of study descriptions and the fact that Webb's discussion focused on how small group research might best be conducted in the future indicated that specialized researchers were clearly her primary audience.

Persons examining Table 2 to discover the key to writing an award-winning review will probably be disappointed. Besides an emphasis on research integration and identification of central issues, the foci and goals that generally define the competition, there is little consistency across the seven papers. In fact, even within the focus and goal categories the papers are not homogeneous. Shavelson, Hubner and Stanton's (1978) paper was primarily a methodological critique of self-concept measures and Thomas' (1982) paper was clearly meant to take issue with the back-to-basics movement. Boruch and Wortman (1979) focused on methodological issues in evaluation research.

The award committees have shown no preference for either neutral or advocacy-type papers, nor for a particular coverage strategy. The organization of papers has primarily been conceptual and the audience primarily specialized researchers. These consistencies, however, are reflective of how often these characteristics appear within the domain of all reviews, as we shall see shortly.

In sum, it appears that reviews of diverse form can be judged to be of the highest quality. The point is important because it underscores the nonjudgmental character of the taxonomy. In fact, the omission of quality criteria from the taxonomy is deliberate. Not to beg this important question, however, I will later return to the problem of what makes a

quality review and how the taxonomy can facilitate quality judgments. First, however, I would like to describe two more applications of the taxonomy. One concerns how reliably the scheme can be applied to describing reviews and the other concerns how often different types of reviews actually appear within the fields of education and psychology.

Assessing the Reliability of Category Placements

While the main purpose of the taxonomy is to catalog the various features of present-day reviews, the scheme would be of added utility if the category labels could be applied to reviews in a reliable manner, that is, with a high degree of consistency across readers. To test whether this was the case, I asked my research assistants, Tom and Ribble, to read and categorize 37 literature reviews in psychology and education. The reviews were chosen from computer printouts of all documents published during the first six months of 1983 that were given the descriptor "literature review" by ERIC or Psychological Abstracts.

Place Table 3 about here

Table 3 presents the intercoder reliability, measured by Cohen's Kappa, and the percent-agreement for each of the six review characteristics. The "First Code" columns relate to the reader's choice of the primary category placement for each characteristic. The "First Plus Second Code" columns define agreement as occurring when both readers nominated a category as either the primary or secondary characteristic of the review. As an example, assume one reader said a review's primary goal was integration and secondary goal was criticism while the other reader said criticism was primary and integration secondary. For the "First Code"

analysis this would be considered a disagreement. For the "First Plus Second Code" analysis this would be considered two agreements.

The results are not very encouraging. For first codes, the Kappa's are unacceptably low. In the case of two categories, goal and coverage, the low reliabilities signify the readers were able to agree on only about half of their judgments. For two other categories, organization and audience, the significantly lower values for Kappa than for percent-agreement indicate that most codes fell into only one category, a conceptual organization and a specialized researcher audience. Kappa, in these instances, adjusts downward the percent-agreement rates to account for what could have been concordance based simply on repeated use of these single categories rather than any "true" discrimination by the readers (see Frick and Simmel, 1978).

The Kappa's for first plus second codes are somewhat more inspiring, based partly on a more even distribution of codes across categories and partly because the definition of agreement was less strict. This latter influence is especially pronounced on the results for the goal category. That is, the two readers categorized many reviews as having the twin goals of integration and identification of central issues. They had difficulty, however, agreeing on which goal was primary and which was secondary.

One might conclude from the reliability data that the taxonomy is poorly defined or does not capture significant distinctions among reviews. I think there are good arguments against such a conclusion. First, the categories are the inductive product of interactions with scholars actively engaged in the reviewing process. Second, few of over one hundred reviewers who were asked to describe their own reviews using the taxonomy have objected to the categories or suggested different ones.

Instead of a lack of clarity in the definitions, I think the low reliabilities reflect the fact that most reviews require multiple codings for several of the categories. For instance, most reviews have more than one focus and goal. Thus, coders are often asked to make subtle distinctions in emphasis. Also reflected in the low reliabilities is a lack of declarative statements on the part of the reviewers concerning what their papers are meant to accomplish and how they were constructed. This lack of information is especially dramatic in the case of the reviewer's coverage strategy, the category which gave the readers the greatest difficulty. To illustrate, Jackson (1980) reported that of 36 reviews randomly chosen from prestigious social science journals, only one gave any indication of the indexes and information retrieval systems used to search the literature and only seven indicated whether they had analyzed the full set of studies or a subset. Such information would clearly help readers discern the intended coverage of a review.

Regardless of the sources of the disagreements, the low reliabilities indicate that if the taxonomy is to be applied by readers, the consensus of multiple readers will be necessary to accomplish a trustworthy categorization of reviews.

A Survey of Literature Review Authors

While readers may have difficulty categorizing reviews, the authors of reviews should find the taxonomy adequately describes their intents and practices. To discover if this was the case and also to obtain some baseline data on how frequently different types of reviews appear in the education and psychology literature, I undertook a survey of recent review authors. The sample for the survey was generated by conducting a computer search of ERIC and PsychInfo in which all documents assigned the descriptor

"literature review" were retrieved that had been published during 1984 and were on-line by December 4, 1984. For ERIC, 168 such documents existed with publication dates through June 1984. For PsychInfo, 100 documents through May 1984 were found. I then excluded 43 documents from ERIC and four from PsychInfo because their abstracts indicated they were primarily annotated bibliographies or project reports. Of the remaining 125 ERIC documents, my research assistants retrieved the addresses of 65 randomly chosen first authors from our university library. Of the 96 PsychInfo documents, 75 first authors were sent questionnaires.[1]

The questionnaire described the taxonomy to authors and asked them to rank order, within each characteristic, those categories that applied to their review while leaving blank those categories that were irrelevant. Of the 140 questionnaires that were mailed, 108, or 77%, were returned completed, seven were returned undelivered, and five were returned with an author comment that their paper was in fact not a review.

The categories provided to authors appeared to adequately capture the vast majority of review characteristics. Authors infrequently made use of the opportunity to provide their own categories to describe their reviews. For example, 12 authors supplied their own description of focus and four of these were simply more specific depictions of categories provided in the taxonomy. No author-offered focus was repeated more than once. Perhaps most troubling to me was the one author who described his work as a "nonempirical, nontheoretical assessment of reality." This response left me wondering why I was studying literature reviews rather than writing such papers myself!

The most curious finding regarding author comments concerned the twelve authors who supplied self-definitions of perspective. I found that

most of these could be reclassified as descriptions of foci or goals. While the percentage of these misunderstandings was small and most of the authors providing self-described perspectives also ranked either the neutral or espousal alternatives, it would be informative to know if the misuse of this category was caused by an unclear definition or by a reluctance on the part of the authors to assert their perspective. I suspect the latter was more often the case.

None of the other characteristics lead more than six percent of reviewers to supply descriptors that were other than specifications of categories already in the taxonomy and no consistency in author-offered descriptors was evident. This indicates that the addition of more categories to the taxonomy is probably unnecessary. In general then, the responses of authors were more encouraging then the reliability of reader codings. Authors apparently felt comfortable describing their reviews in the taxonomy's terms.

Perhaps the best testimony to the taxonomies robustness came from a group of ten reviews abstracted by ERIC that had appeared in a journal called Analytic Chemistry. I debated for some time over whether these papers should be included in the sample, given their somewhat exotic topics, such as "dynamic electrochemistry" and "atomic absorption, atomic fluorescence and flame emission spectometry." I decided to include the papers and only one author returned the survey saying his reply would be inappropriate. All nine other chemistry authors returned the completed questionnaire without comment.

Table 4 presents the descriptive results of the survey. The first two columns list the characteristics and categories. Columns three and four list the percent of respondents who chose each category as a primary or

secondary description of their review. The final column presents the number of reviewers who omitted the characteristic entirely from the description of their paper.

Place Table 4 about here

The responses revealed that about half of all reviews primarily focused on research outcomes, and three of four paid some attention to empirical results. One in five primarily focused on practical applications, and an equal number focused on theory. Only one review in ten took as its primary focus attention to research methods.

The most frequent goal of a review was to critically analyze the relevant literature, with two in five authors saying critical analysis was their primary objective. About one author in four cited formulating general statements and identifying central issues as the primary goal and about one in ten cited resolving conflicts or bridging gaps between theories or ideas as their paramount interest.

The perspective category was dominated by authors who said they took a dispassionate view of the literature (81%) and the organization of most reviews was conceptual (76%).

About two of every three reviewers said they based their conclusions on all of the relevant material and about half of these said all the material was cited in their paper. About one in five reviewers said they used a representative coverage strategy and one in ten a central or pivotal coverage strategy.

About a third of the papers were directed toward specialized scholars, a third toward general scholars, or a third toward practitioners. Policy

makers and the general public were rarely the audience of reviews catalogued by the two abstracting services.

Because of the interest surrounding meta-analysis, I thought it would be informative to determine the percent of reviews that might be considered legitimate candidates for quantitative synthesis. About one reviewer in six (17.6%) claimed their papers primarily focused on research outcomes and had as a goal the formulation of general statements from multiple specific instances. This might be considered a broad definition of a meta-analysis candidate. If we also include in the definition that the author intended to be neutral in perspective and to base conclusions on exhaustive literature coverage then the number of reviews ripe for meta-analysis was one in eight (13%). This finding can be interpreted in two ways. First, advocates of meta-analysis can claim that their techniques are applicable to the largest intersection of review foci and goals. At the same time, however, this type of review represents only a small portion of all literature reviews. The survey, therefore, indicates that other aspects of literature reviewing should not be neglected because of inordinate attention paid to issues surrounding quantitative synthesis.

To discover any relations between the different characteristics of reviews, I performed a correlational and factor analysis of the reviewers' responses. I will only describe some of the results in the most general terms. All the correlations I cite fell between $r = .25$ and $r = .5$ and reached at least the .01 level of significance.[2]

First, reviewers tended to view the characteristics of perspective and organization as containing mutually exclusive categories. This was evidenced by negative intra-characteristic correlations, by limited use of secondary rankings, and by unsolicited comments from respondents. With

regard to coverage, an exhaustive strategy was seen as exclusive of all others. However, use of an exhaustive strategy with both representative and central citations appeared frequently (r = .28), as did the two selective strategies (r = .38).

The focus and goal categories revealed some positive intra-characteristic relations. Reviews that focused on research methods also tended to focus on research outcomes (r = .41) or theories (r = .41). The goal of resolving conflicts in the literature frequently appeared in conjunction with either formulating general statements (r = .41) or bridging theoretical gaps (r = .41). Critical analysis frequently appeared with identifying central issues (r = .28).

With regard to inter-characteristic relations, a focus on research outcomes was associated with the goals of formulating generalities (r = .49) and resolving conflict (r = .36) while writing for an audience of either specialized (r = .29) or general scholars (r = .36). Focusing on methods was associated with critical analysis (r = .29) and identifying central issues (r = .33) as goals, a methodological organization (r = .29), and writing for general scholars (r = .35). A theoretical focus was associated with selectively covering works that were representative of the literature (r = .36) and writing for general scholars (r = .25).

A goal of formulating generalizations was associated with exhaustive coverage but selective citation (representative citation, r = .27; central citation, r = .25) and with an audience of scholars (specialized, r = .28; general, r = .30). Bridging theoretical gaps as a goal covaried with selective citation (representative citation, r = .30; central citation, r = 2.6), a historical organization (r = .34) and either a general scholar (r = .32) or policy-maker (r = .34) audience.

While these associations appear intuitively appealing they should not lead to a conclusion that a small number of similarly-structured prototypes underly most reviews. The factor analysis I performed revealed a first principle component explaining only five percent of the variance and ten factors with eigenvalues greater than one. Beyond the integrative research review, which I estimate accounts for no more than 20% of the review population, no frequently occurring pattern of multiple review characteristics was discernible. Both the descriptive and relational data reveal a body of scholarship, called literature review, that is diverse and held together only by the broadest tenets of secondary analysis and critical synthesis that form the general definition.

Using the Taxonomy to Help Judge the Quality of Reviews

Perhaps the most perplexing question stemming from the increased dependence on literature reviews as a source of information concerns how to distinguish good reviews from bad ones. We have seen that diverse types of reviews exist and there is no reason to believe one type is intrinsically more valuable or valid than another. General discussions of review quality, therefore, will employ criteria of a highly abstract nature, leaving much to the judgment of the individual assessor.

Strike and Posner (1983) suggested that the question of synthesis quality has two parts. The first part involves the intellectual quality and soundness of the synthesis and the second involves its utility. With regard to intellectual quality, Strike and Posner offered three criteria. First:

> A quality synthesis will clarify and resolve, rather than obscure, inconsistencies or tensions between material synthesized. (p 356-357)

Second:

> A quality synthesis will result in a progressive problem shift ... exhibit[ing] such features as increased explanatory and predictive power and expanded empirical content, increased theoretical ability to explain ideas synthesized, expanded scope of application, and an increased capacity to identify and pursue unsolved problems. (p. 357)

And finally:

> A successful synthesis will satisfy the formal criteria for good theories. Such standards as consistency, parsimony, elegance, and fruitfulness characterize a good synthesis. (p. 357)

Strike and Posner's (1983) criteria are indisputable elements of good syntheses. The difficulty in applying them, however, goes beyond the fact that they involve a great deal of subjective judgment. Especially for the criteria of resolving conflict and creating progressive problem shifts, the ability to assess whether a review has performed these functions may take years to develop, since they are dependent on the impact the synthesis has on a field, rather than solely on the intrinsic qualities of the synthesis itself. The third criterion, involving consistency, parsimony, and elegance, strikes closer to the kind of criteria individual readers will apply when they first encounter a review.

There is some empirical evidence that readers of reviews do, in fact, intuitively employ something like Strike and Posner's (1983) third criteria. In a recent study of quality criteria for research reviews (Cooper, in press), I asked fourteen graduate students in education and psychology to read six reviews on the effect of desegregation on black student's achievement. The six reviews had been written simultaneously as

part of a panel sponsored by NIE's Desegregation Studies Team. First, readers were asked to make judgments concerning five aspects of each review: the clarity of problem definition, the exhaustiveness of the covered research, the validity of the reviewers' evaluations of the covered studies, the quality of the synthesis, and the clarity of writing. Factor analyses revealed that a single quality component accounted for between 58% and 83% of the variance in the five dimensions when the analysis was conducted separately for each review. A standardized composite of the five judgments correlated between r = .84 and .96 with the readers' overall judgment of a review's quality. It might be concluded, then, that reviews that were seen as strong on one quality dimension also tended to be seen as strong on others.

A second analysis related the readers' quality judgments to their perceptions of the reviewer's position concerning the effects of desegregation. Somewhat surprisingly, quality judgments were not related to the reviewers' or readers' position on the desegregation issue. Instead, the quality of a review was positively correlated with the readers' confidence in how clearly they could interpret where the reviewer stood on the issue. More interpretability meant higher quality ratings.

The preeminence of presentation factors on quality judgments was further substantiated by open-ended evaluative comments supplied by the readers. In these comments, the readers most often mentioned that a paper was either well or poorly organized. Second most frequently mentioned was writing style, in particular the author's ability or inability to keep the interest of the reader. Third was how well or poorly focused the paper was on the desegregation issue and fourth was how well or poorly the reviewers

used citations to substantiate their claims. All of these criteria appear to be congruent with Strike and Posner's third notion.

In sum, then, the general criteria for good literature reviewing range from the lofty pursuits of resolving conflict and stimulating progressive problem shifts to the rather down-to-earth concern of presenting material effectively enough to get one's point across. Delineating such abstract criteria, however, may not clarify the problem of how to judge quality as much as make potential reviewers wary about pursuing such a complex task.

The taxonomy, I think, can help make evaluating reviews more manageable. Strike and Posner's (1983) second part to assessing quality involved utility. In their words, ". . . useful syntheses will be syntheses which answer the question asked" (p 357-358). This can be translated into two questions involving the six characteristics of reviews. First, do the foci, goals, perspective, coverage, organization and audience of the review form a logical whole? Second, does the review attend to the foci, meet the goals, and employ the expository design the reviewer set for it? I will examine each question in turn.

With regard to the logic of a review, we can ask whether an author has chosen a set of characteristics that are internally consistent. For instance, reviewers who establish the goal of integrating research to form general statements are being inconsistent if they couple this objective with a selective coverage of the literature. Likewise, exhaustive citation of a literature would be counter-productive for a review with the goal of identifying central issues, or for one written for practitioners or policy makers. Obviously, assessing the congruence of matchings could go on. While I would have liked to provide a complete list of what characteristics do and don't fit together, I don't believe such a cataloging is possible.

Some matchings may make sense for certain topic areas but not for others. The point is that having a common, structured scheme for discussing the characteristics of reviews allows assessment and debate of matchings that do occur.

The taxonomy also allows readers to more comprehensively judge whether a review did what it set out to do. An author who claims to have written a review of practices, meant to identify central issues, from a dispassionate perspective, and covering all the relevant literature, provides readers with several self-imposed standards. The key to the use of the taxonomy in this fashion, of course, lies in the willingness of reviewers to state explicitly what they are up to. As we have seen, for certain characteristics such clarity has not prevailed in the past. I would suggest that more important than the creation of any single, uniform scheme for describing reviews is that authors of reviews thoroughly describe the intent and nature of their work, in whatever terms make them comfortable.

The importance of authors describing their intentions and practices recently became personally salient to me when I was asked to be a member of this year's AERA Research Review Award Committee. Being a committee member gave me the opportunity to examine my own process of review evaluation. As I read through the numerous candidates, I found I could not even begin the task until I thought I understood why and how a review was being carried out. Next, I mused over whether the foci, goals, and procedures of the review contained a credible internal logic. These two judgments were minimum criteria. If a review did not meet them I could not give it further consideration. The next step involved a judgment of the complexity or difficulty of the task set by the reviewer. Thus, goals became an

important discriminating criterion. In general, I found linguistic bridge-building and conflict resolution were more lofty pursuits than generalization, criticism, or issue identification, though reviews that attempted to integrate large bodies of literature relating to broad concepts were also deemed very worthwhile. The final criteria concerned how well the authors accomplished the goals they set for themselves. Thus, the evaluation was much like weighting the performance of a dive or gymnastic routine by its degree of difficulty. The conjunction of the last two judgments, I think, was meant to predict which reviews were most likely to achieve Strike and Posner's (1983) criterion of precipitating a progressive problem shift.

If authors make their aims and procedures clear, the process of judging quality not only becomes more feasible, but it can be more objective, as well. Presently, this increased objectivity is evident in the area of integrative research reviews. For example, earlier I defined a meta-analysis as an integrative review of research outcomes, seeking generalities, synthesizing the entire relevant literature in a dispassionate fashion. A review with such characteristics can be held up to a fairly explicit and objective set of standards. Establishing these criteria has occupied this author's attention for several years (see Cooper, 1984b). Some likely candidates appear in Figure 2. Questions that can be asked about integrative research reviews include: (a) do the operations appearing in the literature fit the review's abstract definitions?; (b) is enough attention paid to the methodological details of studies?; (c) was the literature search thorough?; (d) were studies evaluated using explicit and consistent rules?; and (e) were valid procedures used to combine the results of separate studies? Because the

process of evaluating integrative research reviews parallels that of evaluating primary research, the establishment of quality criteria for these reviews is somewhat easier than for other types of syntheses. It is important to recognize that, beyond the general criteria discussed earlier, no set of specific rules will apply to all types of reviews. Each cluster of review characteristics will require a set of distinct standards.

Place Figure 3 about here

Other Uses for the Taxonomy

Before concluding, there are two other important use for the taxonomy. First, many editors of books and journals, directors of information clearinghouses, and funding agencies solicit reviews from particular authors or solicit review proposals. The taxonomy can be used by editors or agencies to communicate to potential authors what they are after. This can be especially helpful if a review is meant to fulfill a particular need or if multiple reviews on the same topic are being solicited and each review is meant to take a different approach. I will shortly be developing for the ERIC clearinghouses just such an application for the taxonomy, along with a means for evaluating whether using the taxonomy to direct authors' efforts enhances readers' perceptions of the utility of the Information Analysis Products sponsored by the Clearinghouses.

Finally, the taxonomy can be used as a framework for graduate education courses in literature reviewing. Students in education and psychology literally can take five or six statistics and methods courses without ever directly addressing the problems and procedures of literature review. This situation is slowly changing. Hopefully, the taxonomy will

facilitate the development of such courses by suggesting a course structure and relevant issues that courses need to address.

Conclusion

I mentioned earlier that this paper was a working document meant to stimulate discussion. My major aims have been to interest others in a topic that has gained increasing salience in education and psychology and to frame some of the questions the topic has engendered. My own involvement with literature reviews began with the circumscribed task of how to best combine the statistical results of independent empirical studies. Shortly thereafter, I came to believe the literature search was at least as, if not more, crucial to the outcome of research reviews than the synthesis process itself. Finally, I saw that research synthesis was only one species of a broad genus. The genus, however, existed without organizing principles. As both consumers and producers of reviews, education and psychology researchers could only benefit from others improving on my "fieldglasses" and "guidebook".

Footnotes

[1] For PsychInfo, every fourth reviewer on the computer printout was not contacted. For ERIC, every fourth and fifth reviewer was passed over. If an address for a first author could not be found, we returned to the beginning of the list and repeated the procedure. Much more difficulty was encountered in locating ERIC authors--the entire listing was exhausted in obtaining the 65 authors sampled. This was because ERIC contains more documents by doctoral candidates and by authors not affiliated with universities. Our primary sources of addresses were: (a) the publication itself; (b) professional organization directories (i.e., APA and AERA); and (c) directories of American University faculty members.

[2] Correlations and factor analyses were performed on data converted to reflect whether or not a category was mentioned by a reviewer, regardless of its ranking. Thus, if a category received any rank it was given a value of 1, if it was omitted it was assigned a value of 0. A second set of analyses that retained the ranking distinctions but treated them as interval rather than ordinal data produced results similar to those described above.

References

Boruch, R. F. and Wortman, P. M. (1979). Implications of educational evaluation for evaluation policy. Review of Research in Education, 7, 309-361.

Cooper, H. (1984a). Literature searching strategies of integrative research reviews. Technical Report #366. Center for Research in Social Behavior, Columbia, MO.

Cooper, H. (1984b). The Integrative Research Review. Beverly Hills: Sage.

Cooper, H. (in press). On the social psychology of using research reviews: The case of desegregation and black achievement. In R. Feldman (Ed.), Social Psychology Applied to Education. Cambridge, England: Cambridge University Press.

Eisner, E. (1983). Anastasia might still be alive, but the monarchy is dead. Educational Researcher, 12, 13-14, 23-24.

ERIC Processing Manual (1982). Phoenix, AZ: Oryx Press.

Frick, T. and Simmel, M. (1978). Observer agreement and reliabilities of classroom observational measures. Review of Educational Research, 48, 157-184.

Garvey, W., and Griffith, B. (1971). Scientific communication: Its role in the conduct of research and creation of knowledge. American Psychologist, 26, 349-361.

Glass, G., McGaw, B., and Smith, M. (1981). Meta-analysis in Social Research. Beverly Hills, CA: Sage.

Jackson, G. (1980). Methods for integrative reviews. Review of Educational Research, 50, 438-460.

Paulston, C. B. (1978). Bilingual/Bicultural education. Review of Research in Education, 6, 186-228.

Phillips, D. (1983). After the wake: Postpositivistic educational thought. Educational Researcher, 12, 4-12.

Price, D. (1965). Networks of scientific papers. Science, 149, 510-515.

Rosenthal, R. (1984). Meta-analytic Procedures for Social Research. Beverly Hills, CA: Sage.

Shavelson, R. J., Hubner, J. J., and Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. Review of Educational Research, 46, 407-441.

Stipek, D. J., and Weisz, J. R. (1981). Perceived personal control and academic achievement. Review of Educational Research, 51, 101-137.

Strike, K., and Posner, G. Types of synthesis and their criteria. In S. Ward and L. Reed (Eds.), Knowledge structure and use: Implications for synthesis and interpretation. Philadelphia: Temple University Press, pp. 343-362.

Taveggia, T. (1974). Resolving research controversy through empirical cumulation. Sociological Methods and Research, 2, 395-407.

Thesaurus of ERIC Descriptors (9th edition). (1982). Phoenix, AZ: Oryx Press.

Thesaurus of Psychological Index Terms (3rd edition). (1982). Washington, DC: American Psychological Association.

Thomas, J. W. (198). Agency and achievement: Self-management and self-regard. Review of Educational Research, 50, 213-230.

Webb, N. W. (1982). Student interaction and learning in small groups. Review of Educational Research, 52, 421-445.

Witkin, H. A., Moore, C. A., Goodenough, D. R., and Cox, P. W. (1977). Review of Educational Research, 47, 1-64.

FIGURE 1

(Figure Caption)

Figure 1. Percentage of documents described as literature reviews in the Psychological Abstracts and ERIC Databases.

Notes: [a]Psychological Abstracts data is based on the percentage of total documents assigned the descriptors "Literature Review" or "Review of the Literature."

[b]ERIC data is based on the percentage of total documents assigned the descriptors "Literature Review" or "Research Reviews." Totals include both RIE(ED) and CIJE(EJ) documents. In 1980, ERIC instituted a document type designation for literature reviews (070). The use of the descriptor "Literature Review" was deemphasized and the descriptor "Research Review" was eliminated.

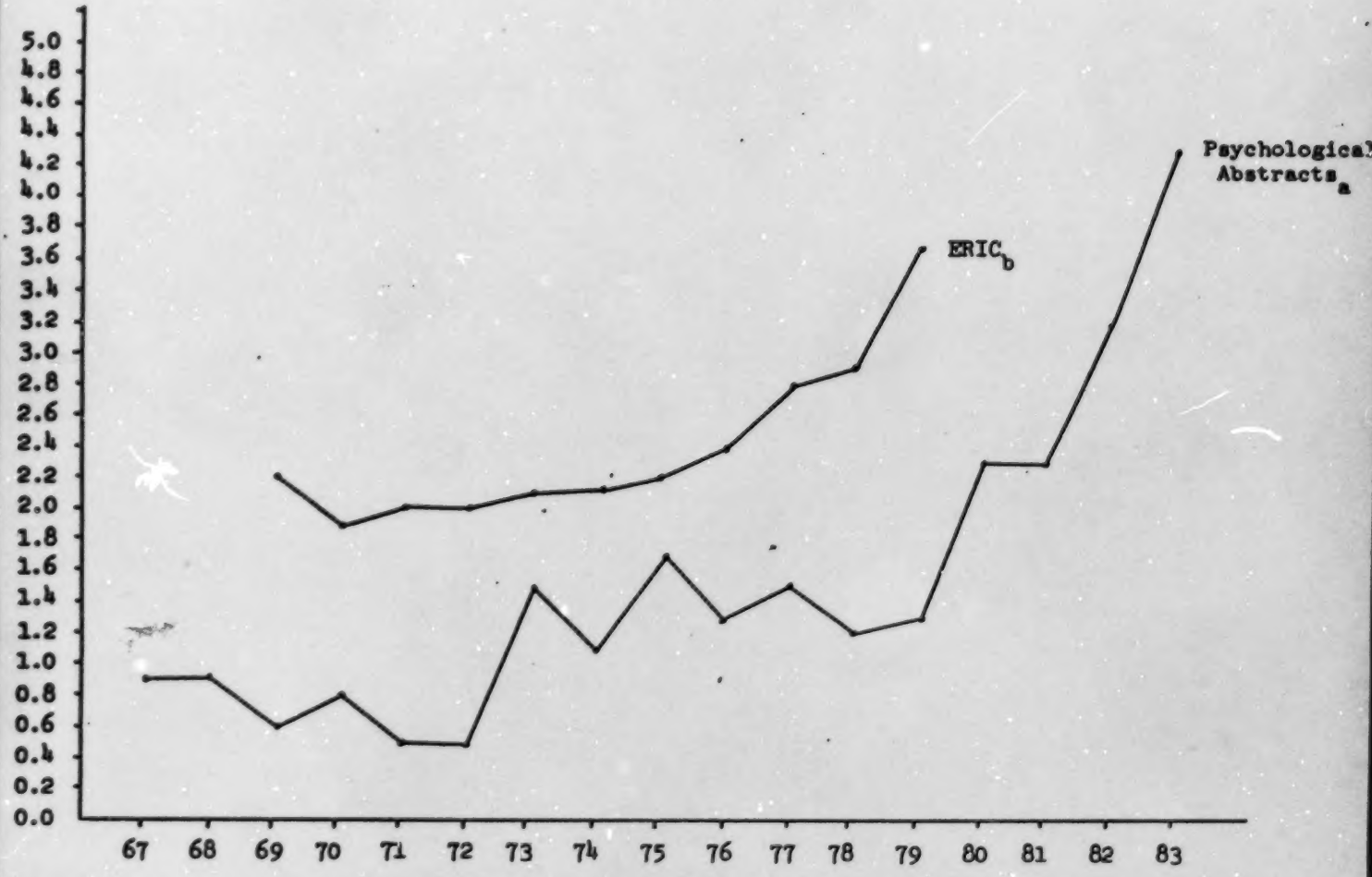
5.0
4.8
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
ERIC b
Psychological Abstracts a

TABLE 1

A Taxonomy of Literature Reviews

Characteristic	Categories
Focus	Research Outcomes Research Methods Theories Practices or Applications
Goal	Integration a) Generalization b) Conflict Resolution c) Linguistic Bridge-building Criticism Identification of Central Issues
Perspective	Neutral Representation Espousal of Position
Coverage	Exhaustive Exhaustive with Selective Citation Representative Central or Pivotal
Organization	Historical Conceptual Methodological
Audience	Specialized Scholars General Scholars Practitioners or Policy Makers General Public

TABLE 2

A Categorization of Reviews Winning AERA's Research Review Award (1978-1984)

Year of Award and First Author

Characteristic	1984 Webb	1983 Stipek	1982 Thomas	1981 Boruch	1980 Paulson	1979 Witkin	1978 Shavelson
Focus	Outcomes Methods	Outcomes (3) Theories (2)	Outcomes (3) Practices (2)	Practices (3)	Outcomes (2) Methods (2)	Outcomes (3) Practices (3)	Methods (3) Theories (1)
Goal	Integration (3) Central Issues (3)	Integration (3) Central Issues (3)	Central Issues (2) Criticism (2)	Central Issues (3) Integration (3)	Central Issues (3) Integration (2)	Central Issues (2) Integration (2)	Criticism (2) Central Issues (2)
Perspective	Neutral (3)	Neutral (3)	Espousal (3)	Neutral (3)	Neutral (3)	Espousal (2)	Espousal (2)
Coverage	Exhaustive (3)	Exhaustive (3)	Selective (3) Represent (3)	Selective (2) Represent (2)	Selective (2) Represent (2)	Represent (2)	Selective (2)
Organization	Conceptual (3)	Conceptual (3) Operations (2)	Conceptual (3)	Conceptual (2)	Conceptual (3) Historical (2)	Conceptual (2) Historical (2)	Operations (2) Conceptual (2)
Audience	Specialized (3)	Specialized (3)	Practitioner (2)	Policy Makers (3)	Specialized (2) Practitioner (2) General (2)	Specialized (2) Practitioner (2)	Specialized (3)

TABLE 3

Reliabilities and Agreement Rates for Two Users of the Taxonomy

	First Code		First Plus Second Code	
	Cohen's K	% Agreement	Cohen's K	% Aggrement
Focus	.48	68%	.55	65%
Goal	.20	48%	.78	86%
Perspective	.53	78%	no second codes	
Coverage	.32	49%	few second codes	
Organization	.23	64%	.45	61%
Audience	.33	84%	.60	73%

FIGURE 2

(Figure Caption)

Figure 2. The Integrative Review Conceptualized as a Research Project

Reprinted with permission from Cooper, H. Scientific guidelines for conducting integrative research reviews. Review of Educational Research, 1982, 52, 291-302. Copyright 1982 by the American Educational Research Association.

Stage Characteristics	Stage of Research				
	Problem Formulation	Data Collection	Data Evaluation	Analysis and Interpretation	Public Presentation
Research Question Asked	What evidence should be included in the review?	What procedures should be used to find relevant evidence?	What retrieved evidence should be included in the review?	What procedures should be used to make inferences about the literature as a whole?	What information should be included in the review report?
Primary Function in Review	Constructing definitions that distinguish relevant from irrelevant studies.	Determining which sources of potentially relevant studies to examine.	Applying criteria to separate "valid" from "invalid" studies.	Synthesizing valid retrieved studies.	Applying editorial criteria to separate important from unimportant information.
Procedural Differences That Create Variation in Review Conclusions	1. Differences in included operational definitions. 2. Differences in operational detail.	Differences in the research contained in sources of information.	1. Differences in quality criteria. 2. Differences in the influence of nonquality criteria.	Differences in rules of inference.	Differences in guidelines for editorial judgment.
Sources of Potential Invalidity in Review Conclusions	1. Narrow concepts might make review conclusions less definitive and robust. 2. Superficial operational detail might obscure interacting variables.	1. Accessed studies might be qualitatively different from the target population of studies. 2. People sampled in accessible studies might be different from target population of people.	1. Nonquality factors might cause improper weighting of study information. 2. Omissions in study reports might make conclusions unreliable.	1. Rules for distinguishing patterns from noise might be inappropriate. 2. Review-based evidence might be used to infer causality.	1. Omission of review procedures might make conclusions irreproducible. 2. Omission of review findings and study procedures might make conclusions obsolete.

TABLE 4

Percentages of Authors Using Various Categories to Describe Their Reviews

Characteristic	Categories	Percentage of Authors Using as Descriptor		
		Primary	Secondary	Omitted
Focus	Research Outcomes	56	12	25
	Research Methods	9	17	47
	Theories	19	23	32
	Practices or Applications	22	23	26
Goal	Integration			
	a) Generalization	23	19	35
	b) Conflict Resolution	7	11	59
	c) Linguistic Bridge-building	8	6	68
	Criticism	42	16	24
	Identification of Central Issues	24	21	34
Perspective	Neutral Representation	81	4	14
	Espousal of Position	18	18	65
Coverage	Exhaustive	37	11	52
	Exhaustive with Representative Citation	21	4	70
	Exhaustive with Central Citation	14	7	75
	Representative	19	6	70
	Central or Pivotal	7	6	78
Organization	Historical	7	9	77
	Conceptual	76	6	17
	Methodological	15	13	70
Audience	Specialized Scholars	39	26	28
	General Scholars	34	23	27
	Practitioners	31	18	28
	Policy Makers	4	5	72
	General Public	2	5	77